Next generation screening using exome sequencing

Genomic Medicine and the Plain Populations of North America
Franklin & Marshall College
Lancaster, PA
July 18, 2013

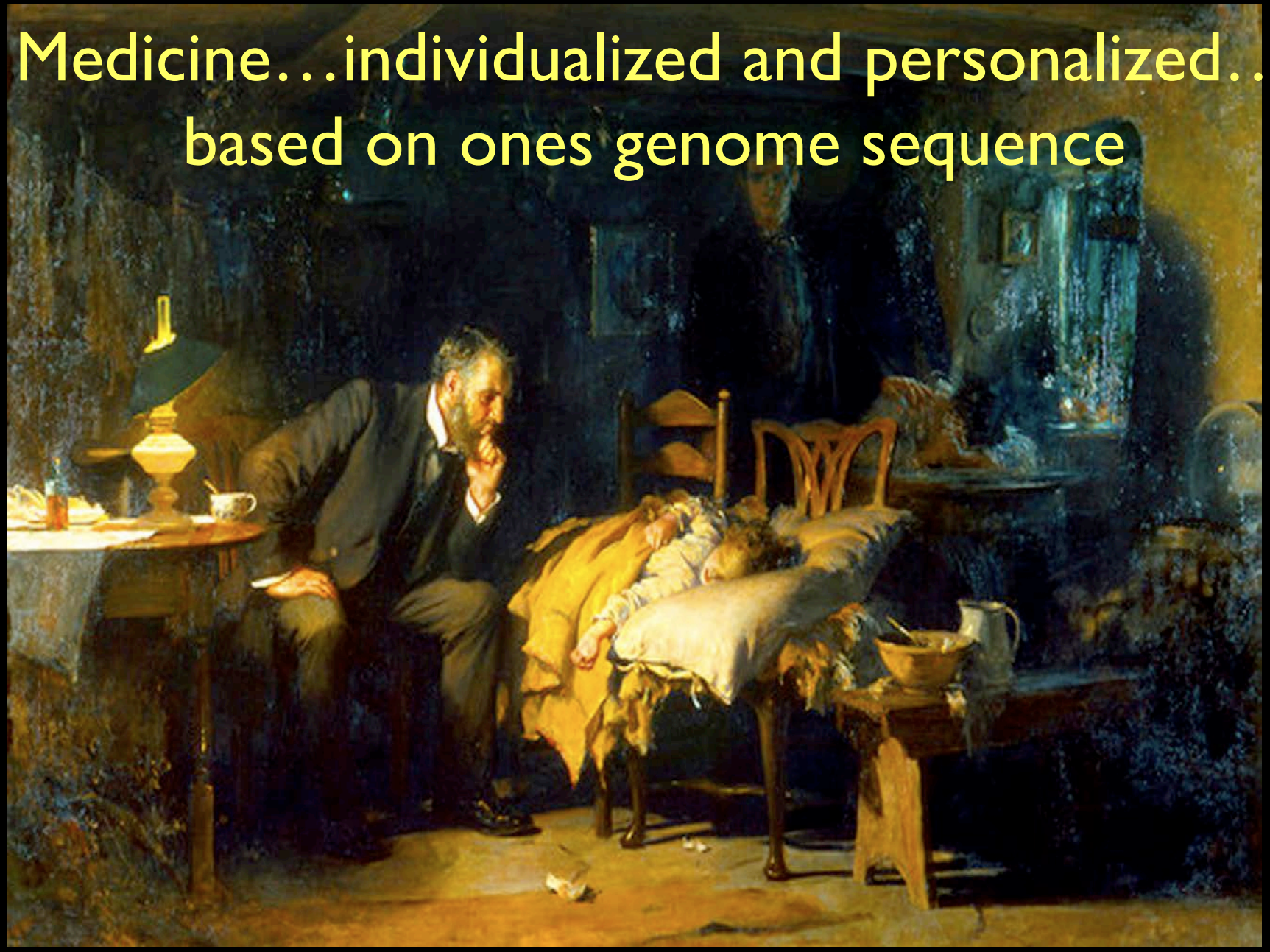
McKUSICK-NATHANS
Institute of
Genetic Medicine

Aravinda Chakravarti, PhD
Center for Complex Disease Genomics
Johns Hopkins University School of Medicine

Disclosures

Aravinda Chakravarti is a paid member of the Scientific Advisory Boards of Affymetrix and Biogen Idec. These potential conflicts of interest are managed by the policies of Johns Hopkins University School of Medicine.

Medicine…individualized and personalized…
based on ones genome sequence



Sir Luke Fildes, The Doctor, 1891

Biology 2.0
Sequence-based Biology (1993-)

- Human Genome Project
- International HapMap Project
- 1000 Genomes Project
- ENCODE Project
- Mendelian Genomics Project

…and many more to come

What progress did we make?

- # mapped disorders: 7,000
- # loci implicated: 5,500
- # disorders with known molecular basis: 4,600

Unfinished business…

- understanding the molecular basis of the remaining 2,000 Mendelian disorders
- understanding the molecular basis of childhood developmental and cognitive disorders

courtesy, OMIM.org, 2012

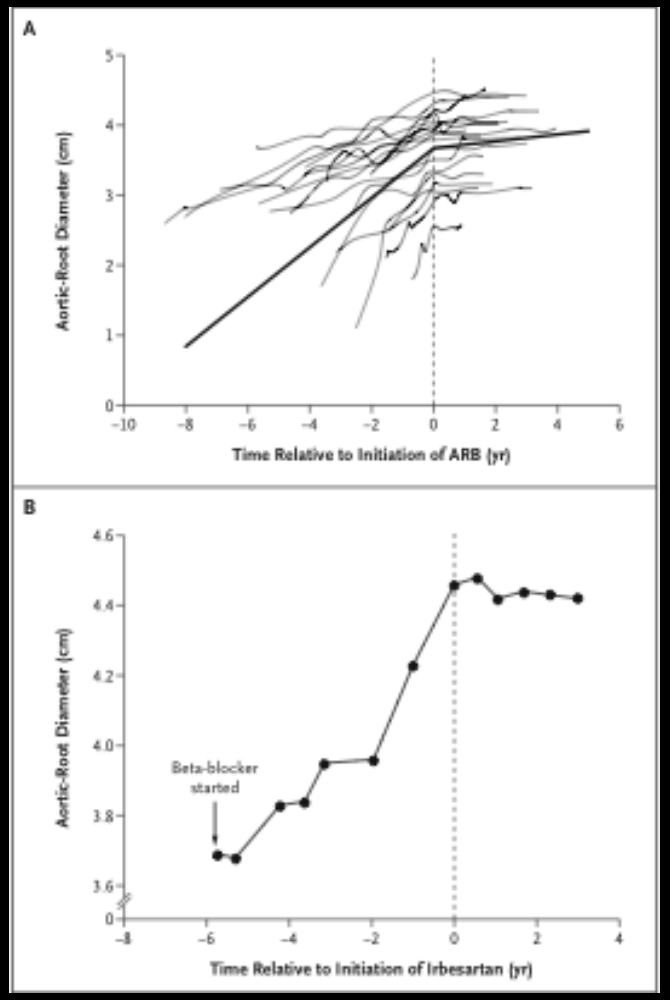
From Mendelian gene identification to informed therapies

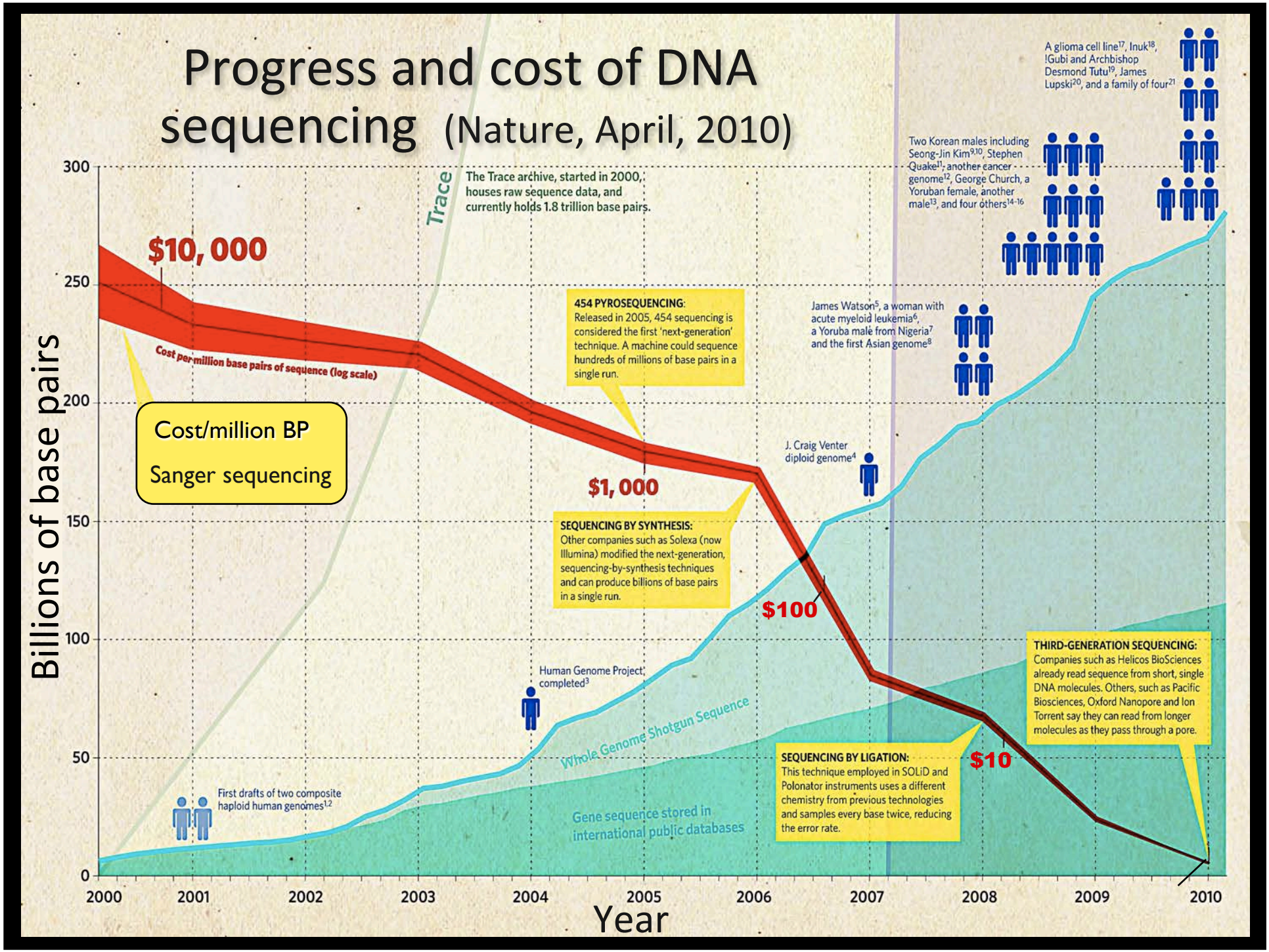
- Marfan syndrome
- Heritable disorder of connective tissue
- Affects the heart & blood vessels, bones/cartilage/ligaments, eyes, lung
- Fibrillin-1 deficiency
- TGF-β activation & losartan
- Other aneurysms?

Angiotensin II Blockade and Aortic-Root Dilation in Marfan's Syndrome

Benjamin S. Brooke, M.D., Jennifer P. Habashi, M.D., Daniel P. Judge, M.D., Nishant Patel, B.A., Bart Loeys, M.D., Ph.D., and Harry C. Dietz III, M.D.

Hal Dietz & colleagues, NEJM, 2008-




Progress and cost of DNA sequencing (Nature, April, 2010)
Billions of base pairs
Year
0
50
100
150
200
250
300
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
Trace
The Trace archive, started in 2000, houses raw sequence data, and currently holds 1.8 trillion base pairs.
$10,000
Cost per million base pairs of sequence (log scale)
Cost/million BP
Sanger sequencing
454 PYROSEQUENCING:
Released in 2005, 454 sequencing is considered the first 'next-generation' technique. A machine could sequence hundreds of millions of base pairs in a single run.
$1,000
SEQUENCING BY SYNTHESIS:
Other companies such as Solexa (now Illumina) modified the next-generation, sequencing-by-synthesis techniques and can produce billions of base pairs in a single run.
$100
SEQUENCING BY LIGATION:
This technique employed in SOLiD and Polonator instruments uses a different chemistry from previous technologies and samples every base twice, reducing the error rate.
$10
THIRD-GENERATION SEQUENCING:
Companies such as Helicos BioSciences already read sequence from short, single DNA molecules. Others, such as Pacific Biosciences, Oxford Nanopore and Ion Torrent say they can read from longer molecules as they pass through a pore.
First drafts of two composite haploid human genomes[1,2]
Human Genome Project completed[3]
J. Craig Venter diploid genome[4]
James Watson[5], a woman with acute myeloid leukemia[6], a Yoruba male from Nigeria[7] and the first Asian genome[8]
Two Korean males including Seong-Jin Kim[9,10], Stephen Quake[11], another cancer genome[12], George Church, a Yoruban female, another male[13], and four others[14-16]
A glioma cell line[17], Inuk[18], !Gubi and Archbishop Desmond Tutu[19], James Lupski[20], and a family of four[21]
Whole Genome Shotgun Sequence
Gene sequence stored in international public databases

Contemporary genomic technologies can effectively sequence and read our genome for Mendelian (single gene) disorders…

this can and will revolutionize Medicine over the next 10 years…

what will this mean for your communities?

The Plain communities have much to teach others about how to effectively use these new technologies to manage and treat genetic dsorders.

The burden of genetic disease in the general population* (1,000 individuals ≤25 yrs)

- Significant genetic component ~ 53 (5.3%)
- Single gene ~ 3.6 (0.4%)
- Autosomal dominant, Autosomal recessive, X-linked recessive ~ 39%, 47%, 14%
- Chromosomal ~1.8
- Multifactorial ~ 46.4
- Unknown but suspected genetic etiology ~ 1.2

*Baird et al, *Am J Hum Genet* 42, 1988

…and this burden is almost surely elevated for recessive disorders in isolated groups

- Founder populations with a limited number of common ancestors and/or the practice of consanguinity
- Incidence increase in large populations depends directly on the degree of consanguinity and inversely on the rarity of the mutation
- Incidence increase in small populations depends inversely on the number of mutation copies

Role of isolated communities in understanding human disease pathophysiology (1950+)

- Victor A. McKusick, The Amish Studies
- J. Perheentupa, Finnish Disease Heritage
- Delineation of new recessive diseases
- Each such entity can illuminate a novel aspect of human disease pathophysiology, and, of course,…
- New routes to disease therapy
- Serial founders and high incidence make gene discovery very efficient

Ellis-van Crevald syndrome (EvC)

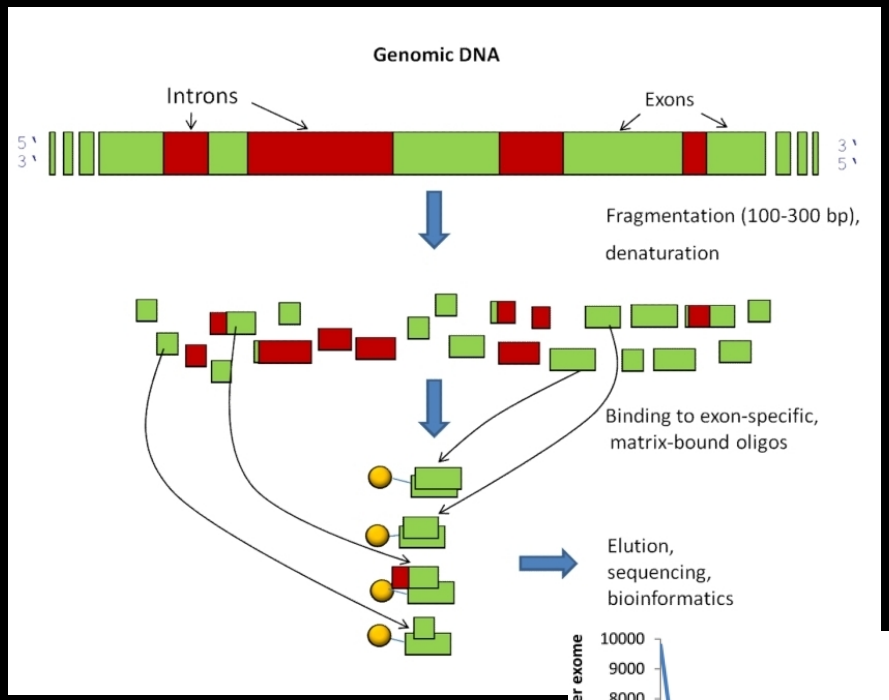
"Amish Madonna and child"

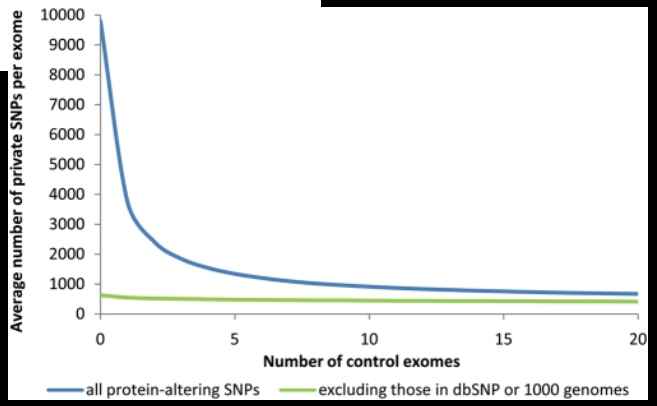
- 52 affected cases
- Inherited from a single ancestor (Samuel Koenig) in 1744
- Incidence ~ 5/1,000 births
- Carrier frequency ~13%
- Single mutation (IVS13+5G>T) in *EVC*
- Developmental regulation of chondrocyte proliferation, hypertrophy and osteoblast differentiation

Exome sequencing identifies the cause of a mendelian disorder

Sarah B Ng[1,10], Kati J Buckingham[2,10], Choli Lee[1], Abigail W Bigham[2], Holly K Tabor[2,3], Karin M Dent[4], Chad D Huff[5], Paul T Shannon[6], Ethylin Wang Jabs[7,8], Deborah A Nickerson[1], Jay Shendure[1] & Michael J Bamshad[1,2,9]

Miller syndrome
Nat Genet
2010





- Capture of coding sequences from blood or tissue DNA
- DNA sequencing
- Alignment of patient's genome to the human reference genome
- Identify all sequence differences
- Identify mutation(s)

Contributions by the Clinic for Special Children

OPEN ACCESS Freely available online

PLoS one

Genetic Mapping and Exome Sequencing Identify Variants Associated with Five Novel Diseases

Erik G. Puffenberger[1,2]*, Robert N. Jinks[2], Carrie Sougnez[3], Kristian Cibulskis[3], Rebecca A. Willert[2], Nathan P. Achilly[2], Ryan P. Cassidy[2], Christopher J. Fiorentini[2], Kory F. Heiken[2], Johnny J. Lawrence[2], Molly H. Mahoney[2], Christopher J. Miller[2], Devika T. Nair[2], Kristin A. Politi[2], Kimberly N. Worcester[2], Roni A. Setton[2], Rosa DiPiazza[2], Eric A. Sherman[4], James T. Eastman[5], Christopher Francklyn[6], Susan Robey-Bond[6], Nicholas L. Rider[1,2,7], Stacey Gabriel[3], D. Holmes Morton[1,2,7], Kevin A. Strauss[1,2,7]

1 Clinic for Special Children, Strasburg, Pennsylvania, United States of America, **2** Department of Biology and Biological Foundations of Behavior Program, Franklin & Marshall College, Lancaster, Pennsylvania, United States of America, **3** The Broad Institute, Boston, Massachusetts, United States of America, **4** Department of Biology, Swarthmore College, Swarthmore, Pennsylvania, United States of America, **5** Department of Pathology and Laboratory Medicine, School of Medicine and Public Health, University of Wisconsin, Madison, Wisconsin, United States of America, **6** College of Medicine, University of Vermont, Burlington, Vermont, United States of America, **7** Lancaster General Hospital, Lancaster, Pennsylvania, United States of America

Abstract

The Clinic for Special Children (CSC) has integrated biochemical and molecular methods into a rural pediatric practice serving Old Order Amish and Mennonite (Plain) children. Among the Plain people, we have used single nucleotide polymorphism (SNP) microarrays to genetically map recessive disorders to large autozygous haplotype blocks (mean = 4.4 Mb) that contain many genes (mean = 79). For some, uninformative mapping or large gene lists preclude disease-gene identification by Sanger sequencing. Seven such conditions were selected for exome sequencing at the Broad Institute; all had been previously mapped at the CSC using low density SNP microarrays coupled with autozygosity and linkage analyses. Using between 1 and 5 patient samples per disorder, we identified sequence variants in the known disease-causing genes *SLC6A3* and *FLVCR1*, and present evidence to strongly support the pathogenicity of variants identified in *TUBGCP6*, *BRAT1*, *SNIP1*, *CRADD*, and *HARS*. Our results reveal the power of coupling new genotyping technologies to population-specific genetic knowledge and robust clinical data.

Citation: Puffenberger EG, Jinks RN, Sougnez C, Cibulskis K, Willert RA, et al. (2012) Genetic Mapping and Exome Sequencing Identify Variants Associated with Five Novel Diseases. PLoS ONE 7(1): e28936. doi:10.1371/journal.pone.0028936

What can exome sequencing detect today?

- Known Mendelian disorders for which the gene is not known;
- Known Mendelian disorders with locus heterogeneity for which known genes ruled out;
- Novel disorders with evidence (mainly pedigree) for Mendelian origin;
- …even suspected Mendelian, mitochondrial and oligogenic (rare) disorders

What can exome sequencing detect today? Mutational types & tissues

- Substitutions;
- Small (<10nt) and large (>5kb) insertions and deletions;
- Balanced chromosome abnormalities using special libraries;
- Any DNA-containing tissue and from blood spots.

What can exome sequencing detect today? Mutations and Functional polymorphisms

- Germ-line and *de novo* disease mutations on autosomes, sex chromosomes and mitochondria;
- Somatic mutations in cancer and other disorders;
- Markers for blood and tissue typing;
- Pharmacogenetic (drug metabolism) polymorphisms;
- Ancestry-specific markers.

If we can identify disease-causing variants at birth in all babies, or for that matter in any asymptomatic individual, then should we do it?

cost vs. benefit?

but who decides?

Value of newborn exome sequencing: Pros

- Technically achievable at reasonable (decreasing) cost, high accuracy and fast turn around time…although the 'implementation' research has not been done;

- Can have high utility to resolve diagnosis for known classes of disorders;

- Identify actionable but unknown disease risks within a community;

- Reduce unnecessary testing, improved patient management and even treatment in many cases (e.g., autism from defects in branched chain amino acid metabolism).

Value of newborn exome sequencing: Cons

- Identify many mutations leading to untreatable disorders (but they wont be untreatable forever…MSUD, GAI);
- Identify many mutations of uncertain risk ("incidentalome");
- Identify genes impacting adult-onset complex disorders (e.g., mental illness, breast cancer);
- Increase anxiety, cost of follow-up, social stigma, loss of employment and loss of health insurance.

The value of exome sequencing in screening:

- Is it appropriate to test for genetic disorders for which we have no treatment/management?

- Is it appropriate not to test for genetic disorders for which we may have treatment/management?

Answering these two questions is the crux of sequence-based newborn testing/screening but this is easier in communities with a high burden of genetic disease from known mutations…value of gene discovery

An ideal

- The Plain communities, through their shared history, represent a common gene pool with many shared genetic mutations and a high risk burden…customized gene screening;

- A mutation of high frequency in one group can often be observed in another;

- Gene discovery across the groups can benefit all, including populations outside;

- Treatment/management strategies in one can analogously benefit other groups.

THANK YOU